Recap

• (Ch 6) Drawing conclusions from a sample of the population
  • Calculating the standard error of the sample mean
  • Constructing confidence intervals around the sample mean

Today

• (Ch 6) Drawing conclusions from a sample of the population
  • Calculating the standard error for other statistics (e.g. sample median)
  • Constructing confidence intervals for other statistics (e.g. sample median)
• (Ch 7) Assessing the significance of evidence against a hypothesis
Calculating standard error of sample mean

• Calculate the unbiased estimate of the population standard deviation

\[
\text{stdunbiased}(\{x\}) = \sqrt{\frac{1}{N-1} \sum_{x_i \in \text{sample}} (x_i - \text{mean}(\{x_i\}))^2}
\]

• The standard error is estimated as

\[
\text{stderr}(\{x\}) = \frac{\text{stdunbiased}(\{x\})}{\sqrt{N}}
\]
Constructing confidence intervals for $N \geq 30$

- 68% confidence interval
  \[ [\text{mean}(\{x\}) - \text{stderr}(\{x\}), \text{mean}(\{x\}) + \text{stderr}(\{x\})] \]

- 95% confidence interval
  \[ [\text{mean}(\{x\}) - (2)\text{stderr}(\{x\}), \text{mean}(\{x\}) + (2)\text{stderr}(\{x\})] \]

- 99% confidence interval
  \[ [\text{mean}(\{x\}) - (3)\text{stderr}(\{x\}), \text{mean}(\{x\}) + (3)\text{stderr}(\{x\})] \]
What does a 99% confidence interval mean?

• For 99% of the samples, if you construct the confidence interval in this way, the population mean will lie within the interval.

• It does not mean that the population mean lies in the interval with probability 99%.

³ This does not make sense in a frequentist point-of-view.
Standard error: midterm grading example

The realized sample of scores is \( \{120, 130, 140, 140, 150\} \) with \( N = 5 \)

mean(\( \{120, 130, 140, 140, 150\} \)) = 136

\[
\sqrt{\frac{(120 - 136)^2 + (130 - 136)^2 + 2(140 - 136)^2 + (150 - 136)^2}{5 - 1}} = 11.4
\]

stdunbiased(\( \{x\} \))

\[
\text{stderr}(\{x\}) = \frac{11.4}{\sqrt{5}} = 5.1
\]

So we estimate the population mean as 136 with standard error 5.1
Standard error: election polling example

\[
\text{stdunbiased}(\{x\}) = \sqrt{\frac{715(0.49)(1 - 0.49)^2 + 715(0.51)(0 - 0.49)^2}{715 - 1}} = 0.50
\]

\[
\text{stderr}(\{x\}) = \frac{0.5}{\sqrt{715}} = 0.019
\]

So we estimate the population mean as 49% with standard error 19%.
Confidence interval: election polling example

<table>
<thead>
<tr>
<th></th>
<th>DATES</th>
<th>POLLSTER</th>
<th>SAMPLE</th>
<th>RESULT</th>
<th>NET RESULT</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. House</td>
<td>* IL-12 SEP 26-27</td>
<td>DCCC Targeting Team*</td>
<td>574 LV</td>
<td>Kelly 44% 48% Bost</td>
<td>Bost +2</td>
</tr>
<tr>
<td></td>
<td>* IL-12 SEP 26-27</td>
<td>DCCC Targeting Team*</td>
<td>574 LV</td>
<td>Kelly 41% 42% Bost</td>
<td>Bost +1</td>
</tr>
<tr>
<td>Governor</td>
<td>* Ill. SEP 24-29</td>
<td>Southern Illinois University</td>
<td>715 LV</td>
<td>Pritzker 49% 27% Rauner</td>
<td>Pritzker +22</td>
</tr>
</tbody>
</table>

Source: fivethirtyeight.com

We estimated the population mean as 49% with standard error 1.9%

The 99% confidence interval for Pritzker’s vote percentage is

\[ [49\% - (3)1.9\%, 49\% + (3)1.9\%] = [43.3\%, 54.7\%] \]
Confidence intervals for other statistics

• The **bootstrap** is a method to construct confidence intervals for other statistics (besides the sample mean) for which we cannot derive analytical expressions

• Bootstrapping a confidence interval for the sample median given a sample of $N$ items
  • Create a bootstrap replicate by sampling $N$ items from the original sample uniformly and **with replacement**
  • Calculate the sample median of the bootstrap replicate
  • Repeat the two steps above a large number of times
  • Plot the histogram of the sample medians and construct a confidence interval
Bootstrap median: midterm grading example

The realized sample of scores is \{120, 130, 140, 140, 150\} with \(N = 5\)

\[
\text{median}\left(\{120, 130, 140, 140, 150\}\right) = 140
\]

<table>
<thead>
<tr>
<th>Bootstrap replicate</th>
<th>Sample median</th>
</tr>
</thead>
<tbody>
<tr>
<td>{150, 150, 130, 130, 140}</td>
<td>\textbf{140}</td>
</tr>
<tr>
<td>{120, 120, 140, 140, 120}</td>
<td>\textbf{120}</td>
</tr>
<tr>
<td>{130, 140, 140, 150, 140}</td>
<td>\textbf{140}</td>
</tr>
<tr>
<td>{150, 140, 140, 150, 140}</td>
<td>\textbf{140}</td>
</tr>
<tr>
<td>{150, 120, 130, 130, 140}</td>
<td>\textbf{130}</td>
</tr>
</tbody>
</table>
Bootstrap median: midterm grading example

Repeat a total of 10000 times and plot a histogram of sample medians

Histogram of sample medians from 10000 bootstrap replicates

90% confidence interval for sample median
The scientific method

• Form a hypothesis about some phenomenon

• Design an experiment and collect data

• Reject the hypothesis if it is contradicted by the data
Hypothesis testing: election polling example

Hypothesis: Pritzker’s vote percentage is 53%

Experiment

<table>
<thead>
<tr>
<th>Poll</th>
<th>Dates</th>
<th>Pollster</th>
<th>Sample</th>
<th>Result</th>
<th>Net Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. House</td>
<td>* IL-12</td>
<td>SEP 26-27</td>
<td>DCCC Targeting Team*</td>
<td>Kelly 44% 40% Bost</td>
<td>Bost +2</td>
</tr>
<tr>
<td>U.S. House</td>
<td>* IL-12</td>
<td>SEP 26-27</td>
<td>DCCC Targeting Team*</td>
<td>Kelly 41% 42% Bost</td>
<td>Bost +1</td>
</tr>
<tr>
<td>Governor</td>
<td>* Ill.</td>
<td>SEP 24-29</td>
<td>Southern Illinois University</td>
<td>Pritzker 49% 27% Rauner</td>
<td>Pritzker +22</td>
</tr>
</tbody>
</table>

Should we reject the hypothesis based on this data?
Fraction of “less extreme” samples

• Assuming that the hypothesis is true, what fraction of samples would have had sample means less extreme than what we observed?

\[
0.49 \quad 0.53
\]

\[
\frac{(\text{sample mean}) - (\text{hypothesised value})}{\text{standard error}}
\]

\[
0.019
\]

• Define a test statistic \( g \)

• If \( N \geq 30 \), we can say \( g \) comes from a standard normal distribution

• So, the fraction of “less extreme” samples

\[
f = \frac{1}{\sqrt{2\pi}} \int_{-|g|}^{|g|} \exp\left(-\frac{x^2}{2}\right) dx
\]
P-value: fraction of “more extreme” samples

• It is conventional in science to report the p-value of an experiment

\[ p = 1 - f = 1 - \frac{1}{\sqrt{2\pi}} \int_{-|g|}^{|g|} \exp \left( -\frac{x^2}{2} \right) dx \]

• So, a p-value is the fraction of samples that would have had sample means more extreme than what we observed, assuming that the hypothesis is true

• By convention, if the p-value < 0.05, we should reject the hypothesis
P-value: election polling example

• Hypothesis: Pritzker’s vote percentage is 53%

• Recall that we calculated sample mean 49% and standard error 1.9%

• So the test statistic $g = \frac{49-53}{1.9} = -2.11$

• And the p-value tells us to reject the hypothesis

$$p = 1 - \frac{1}{\sqrt{2\pi}} \int_{-2.11}^{2.11} \exp \left( -\frac{x^2}{2} \right) dx = 0.035 < 0.05$$
The use and misuse of p-values

• P-values in scientific practice
  • Scientists are usually trying to reject the null hypothesis, the hypothesis that there is no special phenomenon and the data are just random noise
  • So $p < 0.05$ means there may be an interesting phenomenon and it has been the standard for publication in many fields

• What’s wrong with using p-values in this way?
  • Rejecting the null hypothesis doesn’t mean that the proposed alternative hypothesis is true
  • $p < 0.05$ is arbitrary and gives a 1-in-20 chance of false positives
  • It encourages p-value hacking and has contributed to the replication crisis